Intelligent Diagnostic Monitoring Using Trend Templates

Ira J. Haimowitz, Ph.D., Corporate Research and Development,
General Electric Company, Schenectady, NY*

ABSTRACT

In previous work we have defined our *trend template* epistemology for clinically significant trends and we have illustrated and tested a program *TrenDx* that monitors time-ordered process data by matching the data to trend templates. Our initial application domain was pediatric growth monitoring. In continuing work we have explored monitoring hemodynamic and respiratory parameters of intensive care unit patients. This application has highlighted the needs for advances in our representation and monitoring algorithms. In particular, we have added reasoning with uncertainty to the trend template epistemology, and a new control structure allowing numerical ranking of competing trend templates. Furthermore, intelligent monitoring in any medical domain requires a coherent framework for diagnostic monitoring. In this paper we show how TrenDx can be extended to a framework including sending alarms, changing clinical context, and filtering data streams. [1]

INTRODUCTION

Trend Templates and TrenDx

In previous work [1, 2] we have defined our *trend template* epistemology for clinically significant trends, consisting of *landmark points* representing events in a process and *intervals* representing phases of that process. Below is part of a trend template for normal pediatric growth in boys. This partial template includes two landmark points, representing birth and the onset of puberty, which is at temporal distance 10 to 13 years from birth. The two intervals, representing periods of a child establishing height and weight centiles and of pre-pubertal growth, also have temporal uncertainty as indicated.

Notice also that the two intervals are represented as *consecutive phases*, so that the endpoint of the first interval is infinitesimally before the begin point of the second interval. Also attached to each interval are *value constraints* that restrict the value of relevant parameters that occur in data during the intervals. In particular, these value constraints place limits on the Z-

scores (number of standard deviations from the population mean) of the normal growth patient's heights and weights.
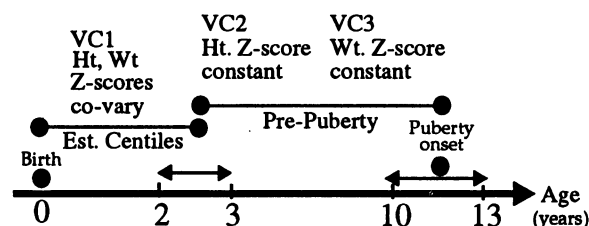

Figure 1 Partial trend template for normal pediatric growth.

We have illustrated and tested a program *TrenDx* that monitors time-ordered process data by matching the data to trend templates. The program branches to consider alternate temporal chronologies of how the process data has evolved.

Previous Work in Pediatric Growth Monitoring

General pediatricians that monitor their patients' growth suffer from data overload in that they have insufficient time per patient to investigate the entire time-series of growth data (heights, weights, and bone ages). Our goal in that domain was to develop a screening tool for general pediatricians that could examine growth chart data and suggest a possible disorder.

As described in [2] we conducted a preliminary clinical trial using 30 consecutive growth records from the endocrinology clinic at Boston Children's Hospital. The results were encouraging in that TrenDx showed promise in reaching the same diagnosis as a panel of experts, at a time no later than the experts, in most of the cases. The trial was also useful in uncovering some representational issues that needed further research. We are currently planning a larger scale trial using hundreds of clinical cases where will we compare TrenDx monitoring performance to humans of various expertise: medical students, general pediatricians, and pediatric endocrinologists.

Intensive Care Unit Monitoring

We have attempted to apply our approach to the domain of intensive care unit (ICU) monitoring. Here there is also data overload: eight or more patients are in an ICU, and each patient is monitored with dozens of hemodynamic and respiratory variables sampled

several times per minute. It is impossible for nurses to steadily monitor even minutes worth of continuous data from an individual patient. Our goal in ICU monitoring is developing context-sensitive monitors whose use will significantly reduce the high false positive rates typically produced with single-variable threshold monitors. TrenDx can potentially monitor the adequacy or failure of external interventions on ICU patients, and the normality or abnormality of physiological mechanisms in these patients. Interventions and mechanisms may display a characteristic multivariate pattern over several phases. An automated monitor must apply specific filters and value constraints appropriate for each phase of the intervention or mechanism.

In Figure 2 are one hour of ICU data from an 8 month old girl with adult respiratory distress syndrome [3]. Four signals are plotted from 12:00 a.m. to 1:00 a.m.: heart rate taken from the electrocardiogram (ECG), mean arterial blood pressure, oxygen saturation, and fraction of inspired oxygen (FIO_2). Data were compressed by reporting values only upon changes. Usually, the patient received oxygen via the ventilator, FIO_2 at 50%. Approximately once every two hours, the patient was ventilated by the nurse squeezing a hand bag filled with 100% oxygen, so that a bronchodilator could be delivered in aerosol form. One handbagging session was from 12:22 a.m. until 12:31 a.m. As illustrated in the figure, the change to hand-bagging was marked by an immediate rise of FIO_2 from 50% to 100%, remaining at 100% during hand bagging. Within a minute after hand-bagging began, O_2 saturation rose sharply to 100%. These two changes are expected in such a handbagging session. During such hand-bagging it is preferable that the patient's hemodynamics remain stable. However, in this patient mean arterial blood pressure dropped from about 12:26 a.m. to 12:31 a.m, and ECG-measured heart rate rose steadily from

approximately 12:27 until 12:30 a.m. These two changes are not usually expected. This pattern in these four parameters occurred during each of the six hand-bagging sessions for this patient over a twelve-hour period.

One possible explanation for this hemodynamic fault is that the oxygen handbagging increased pressure in the chest cavity. This could have depressed the patient's vena cava and compromised her venous return to the heart, resulting in the falling blood pressure. The heart rate increase may have been a normal baroreceptor reflex to the falling heart rate. Whatever the explanation, this hemodynamic fault is worthy of a clinician's attention.

Framework for Intelligent Diagnostic Monitoring

In order to achieve robust monitoring performance in multiple domains, we have extended our trend detection algorithm TrenDx to a broader framework for intelligent diagnostic monitoring. This framework includes a means of representing significant multivariate trends with multiple phases, and methods of detecting those trends from data. Also included are means for generating reliable alarms, displaying and explaining significant trends, and changing the clinical context. More complete details of the framework are in the author's dissertation [4].

REPRESENTING SIGNIFICANT TRENDS

Regression-Based Trend Templates

To advance our original work we needed a trend representation with robust matching, and allowing ranking of competing trend hypotheses. We modified trend template value constraints to be parameterized statistical models describing variation in data assigned to an interval. More precisely, let hyp be a TrenDx hypothesis consisting of a trend template TT; *hyp* assigns data to the intervals of TT. Let *I* be an interval of TT and let
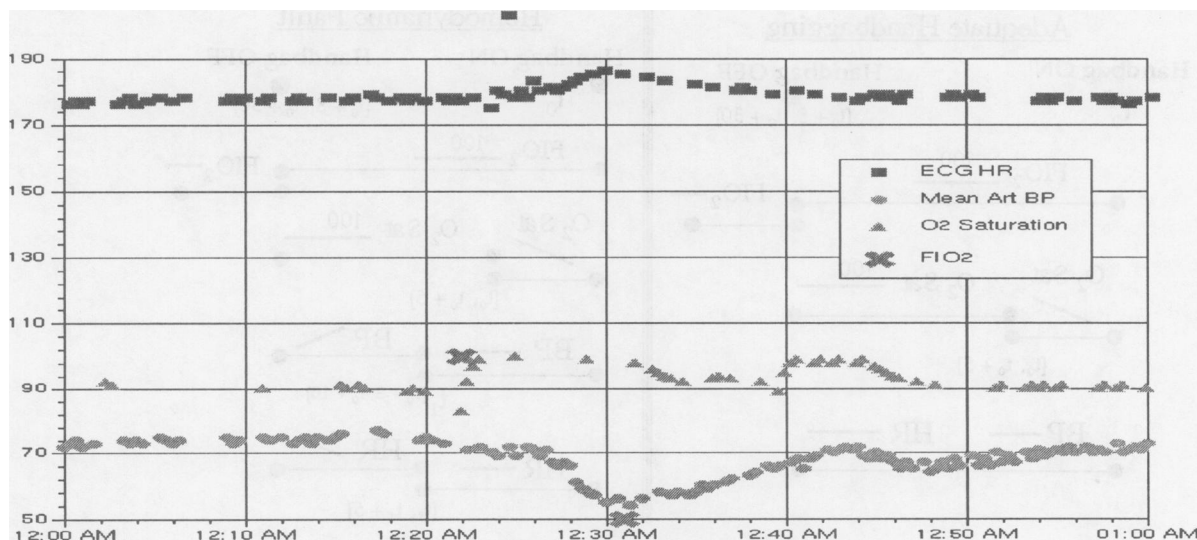

Figure 2: One hour of four signals plotted from an intensive care unit patient.

D(I, hyp) be the data assigned to I in hyp. Each value constraint consists of two main components:

    1.a function F that maps the data D(*I, hyp*) to a time-indexed real-valued sequence $\{Y_t\}$.

    2.a linear regression model describing the pattern of $\{Y_t\}$.

The primary set of linear regression models used in this research are polynomials of degree 0, 1, and 2, with qualitative or quantitative constraints on some subset (perhaps empty) of the polynomial coefficients. A *qualitative constraint* is a member of the set {+, -}, representing that the parameter is positive or negative.[2] A *quantitative constraint* is either a single numerical value or a numerical range [min max] of values. There are seven qualitatively distinct elementary regression models used in value constraints. These seven models are sufficient to roughly distinguish between different types of behaviors.

Constant models with quantitative parameter constraints can be used to represent steady states. An interval of normal human temperature may constrain temperature to be constant at 37 degrees Celsius. A constant model without a numerical estimate represents quiescence at an unknown level.

Linear models with quantitative slope constraints can help distinguish clinically distinct trends. For example, blood pressure loss due to handbagging may have a slope in the range -1 mm Hg to -3 mm Hg per minute, whereas blood pressure loss due to internal hemorrhaging may have a slope in the range -10 mm Hg to -20 mm Hg per minute. Linear models with

---

2. In this research a parameter estimate of 0 is considered a quantitative estimate.

qualitative slopes constraints can roughly distinguish responses.

Quadratic models with qualitative constraints are useful for representing trends having a sharp increase or decrease followed by a stabilization. They can better fit data showing a nonlinear response than can a linear model. When qualitative or quantitative constraints on quadratic coefficients are not derivable, a knowledge engineer may better characterize a quadratic trend with qualitative constraints on the first and second derivatives.

Monitor Sets

A *monitor set* is a set of trend templates forming a clinical context. The trend templates within a monitor set are viewed as a partition of trends that may occur in a particular clinical context. The members of a monitor set are concurrently matched against the same patient data by TrenDx.

In a diagnostic setting one trend template within a monitor set is the *expected* or *normal model*; the other trend templates are *fault models*. The fault models are those that if matched well warrant attention by the person or system observing the device.

Monitor Set for Oxygen Handbagging

In Figure 3 are two trend templates comprising a monitor set of patient response to 100% oxygen handbagging. One trend is expected and the other suggests a fault.

The hemodynamic fault trend template consists of eight intervals. The changes in four parameters are each represented in a pair of intervals. Temporal relations between these intervals establish a pattern that is fairly specific to this particular population response. The top two intervals denote that, during handbagging, the fraction of inspired oxygen (FIO_2) remains con-
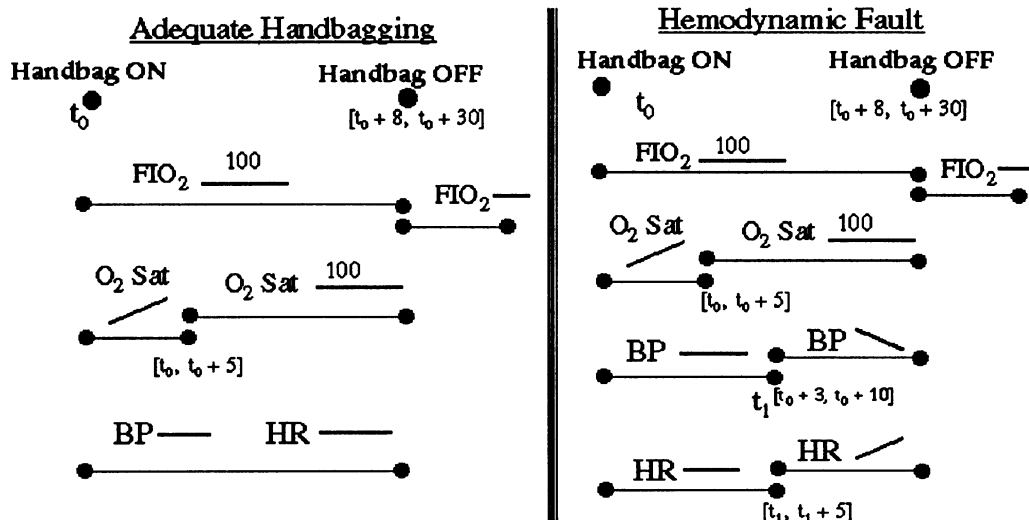

Figure 3: Competing trend templates for oxygen handbagging.

704

stant at 100 percent, and that for three minutes after handbagging, FIO_2 is constant at some unspecified value. Given the time of Handbag On for a particular patient, the trend detection program TrenDx can use these two intervals to estimate that patient's time of Handbag Off. TrenDx estimates based on when FIO_2 has changed from 100 percent. The next two intervals describe the rise and stabilization of oxygen saturation of hemoglobin. During saturating Hb, O_2 saturation is linear and increasing; during saturated Hb, O_2 saturation is constant at 100 percent. The two other pairs of intervals constrain the responses of blood pressure and heart rate. Each parameter first has a steady phase, beginning at the same time as Handbag On. During these phases, both parameters are constant. A second phase of decreasing BP, beginning 3 to 10 minutes after Handbag On, constrains BP to be linear and decreasing. This phase ends at Handbag Off. A second phase of increasing HR, beginning 0 to 5 minutes after the begin point of decreasing BP constrains HR to be linear and increasing. The temporal relations between these intervals insure that as TrenDx matches process data early in the template, the program constrains the expected match to data in the future.

The trend template for adequate handbagging contains the same landmark points and the same response intervals for FIO_2 and O_2 Sat. The adequate handbagging trend template differs in its trends for BP and HR, both of which are constrained to be constant in a single interval whose length is the handbagging period.

MATCHING DATA TO TREND TEMPLATES

The trend diagnosis program TrenDx matches patient data to the regression-based trend templates in each monitor set. TrenDx instantiates the trend templates for a particular patient by anchoring a landmark point of a trend template to an event in the patient history. In this ICU example, the landmark point handbag on is anchored to the time of a special datum generated by a switch on the ventilator noting that the ventilator no longer supplies oxygen. In principle instantiation could also proceed via the results of a strong match to a preliminary trend template.

The goodness of fit of value constraint vc for the hypothesis hyp, denoted by Fit(vc, hyp), is the mean absolute percentage error (MAPE) between sequence values $\{Y_t\}$ and regression model estimations $\{\hat{Y}_t\}$:

$$Fit(vc, hyp) = \frac{\sum_t \left| \frac{\hat{Y}_t - Y_t}{Y_t} \right|}{N - p} \quad \text{(EQ 1)}$$

where N is the number of values within the interval, and p is the number of parameters that are estimated. MAPE is particularly useful for comparing the good-

ness of fit between models of different variables of possibly different measurement scales.

The goodness of fit of a hypothesis to the data assigned to the intervals of its trend template is a weighted average of the fits to the individual value constraints. The weights may be defined by experts; by default the weights are the (N-p) used as the denominators of value constraint scores. For each trend template, which has temporal uncertainty, TrenDx optimizes over all temporal distances to find the best matching hypothesis to the data.

TrenDx matched the trend templates in Figure 3 to four signals of data between 12:22 and 12:31 a.m. during which the patient received oxygen via a handbag. Several intensive care unit physicians agreed the patient was experiencing some hemodynamic fault during this period. TrenDx also matched the same monitor set to five other periods of handbagging in this patient during the same day. The results were similar enough that we only show results of matching to the first handbagging session.

Figure 4 shows the four signals of ICU data and results of TrenDx matching to this data. Note that the outlying heart rate between 12:24 a.m. and 12:25 a.m. caused a jump in the match scores to both trend templates. Had that outlier been removed or smoothed by filtering, the error scores for both trends would have remained lower.

The best matches of each trend template stay close in score until 12:27:21 a.m., when the percentage error for the adequate handbagging trend template rises while that for the compromised venous return stays level. The difference between these two, plotted at the bottom of the graph, rises steadily for the duration of the handbagging episode. This difference, if judged significant, may be used as a means for sending an alarm.

JUDGING TREND SIGNIFICANCE

Generally, TrenDx matches to a monitor set by computing for each time slice of data the best matching score for the normal trend template and for each competing fault. Throughout this section we denote the sequence of best scores for the normal template $\{TT_n\}$ and the sequence for each fault template $\{TT_{fi}\}$. High values in these sequences indicate a poor match to data. We denote by $\{TT_n - TT_{fi}\}$ the sequence of score differences some fault model and the normal model. High values in this sequence indicate that the fault model matches better than the normal model.

We now must devise a scheme for answering the following questions:

- When has the normal model become a significantly poor match to the data to require attention?
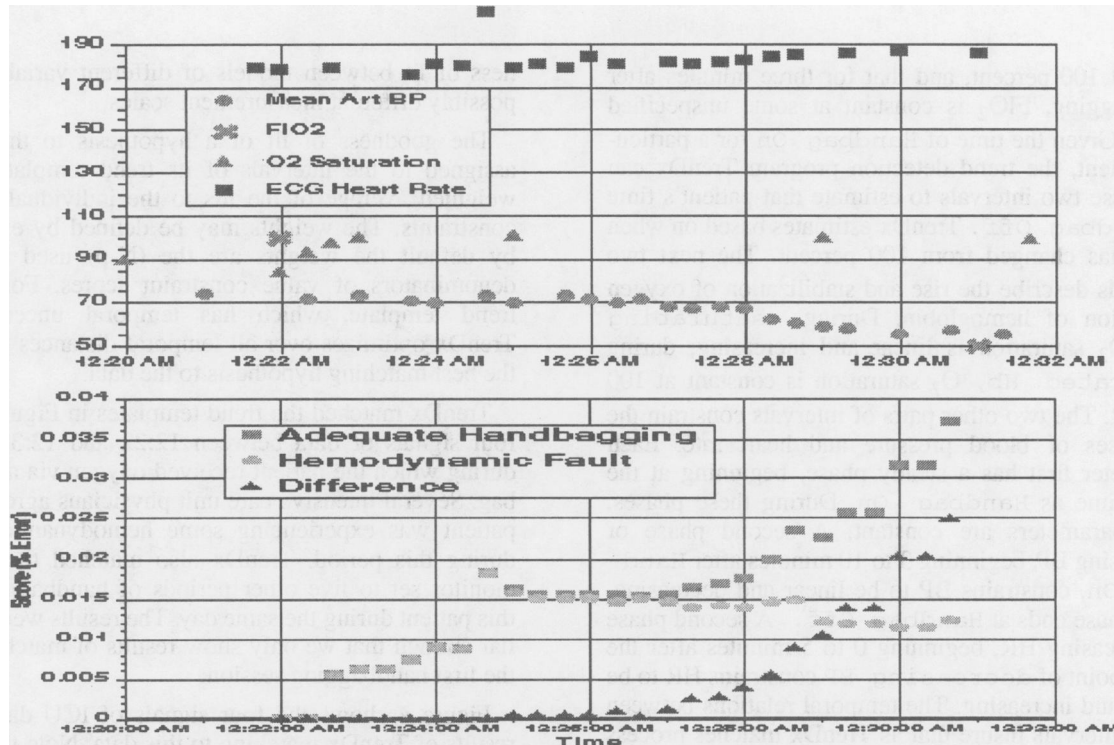 The answer is a property of the sequence $\{TT_n\}$.

705

Figure 4: TrenDx results of monitoring ICU data during handbagging.

- When is a fault model a significantly better match to the data than the normal model?

  The answer is a property of the sequence $\{TT_n - TT_{fi}\}$.

There are many possible means to answer these questions. In fact, trying to answer them uncovers another monitoring problem. Notice that each of the TrenDx result plots in Figure 4 is in fact a time-ordered sequence of data, and we wish to detect particular types of trends in these sequences. Does this mean we are back to square one of diagnostic monitoring? Not really, for we have significantly reduced the complexity of our monitoring task in two respects, First, we no longer must represent the *relevant diagnostic categories*, which requires extensive domain knowledge. Instead, we must represent the *relevant monitoring strategies*, which depend less on the progress of disorders but more on the monitoring environment. Second, we have also reduced the dimension of the relevant temporal patterns we seek. No longer are we finding multivariate patterns measurements, but instead we seek univariate patterns in either the time-sequence of best scores $\{TT_n\}$, or the time-sequence of difference in best scores $\{TT_n - TT_{fi}\}$, for some fault model.

In principle, we may use any univariate trend detection scheme available, including TrenDx. We illustrate a few straightforward methods, based on thresholds and accumulation of differences over time.

Thresholds

The simplest method of determining whether the score for of $\{TT_n\}$ or $\{TT_n - TT_{fi}\}$ is significant is establishing a threshold over which the matching score triggers an alarm. A threshold for a high score of $\{TT_n\}$, denoted by $TT_n{}^*$, is best determined by experience from training cases. Let $\overline{TTn}$ denote the average value of $\{TT_n\}$ over time for a normal patient case (contiguous data set). We presume $\overline{TTn}$ is normally distributed.[3] We recommend a two-stage supervised learning procedure. In the first stage, run a set of normal cases through TrenDx matching to the normal trend template and compute unbiased estimates of the mean and standard deviation of $\overline{TTn}$. In the second stage, run both normal and abnormal cases through TrenDx using different levels of $TT_n{}^*$ for sounding an alarm. Select a value of $TT_n{}^*$ producing satisfactory sensitivity and specificity. If the normality assumption was a close approximation, then choosing $TT_n{}^*$ as two estimated standard deviations above the mean of $\overline{TTn}$ will yield a sensitivity of better than 0.95.

A significant threshold for $\{TT_n - TT_{fi}\}$ may be learned similarly: first estimate the distribution using

---

3. This presumption is based on intuition rather than a statistical result. Because $\{TT_n\}$ is non-negative, there may be some skew. With sufficient cases one can test an empirical distribution for normality [5, page 322].

706

faulty cases, then optimize sensitivity and specificity from normal and faulty cases. One may also find a threshold by relying on intuition, based on a TrenDx score giving a mean percentage error in explaining the data. $\{TT_n - TT_{fi}\}$ being above a threshold p means that the normal trend template is p percent more erroneous than the fault model. One may also consider judging a faulty trend as significant using a threshold for the *percentage better match* of the fault model:

$$\frac{TT_n - TT_f}{TT_n} = 1 - \frac{TT_f}{TT_n} \ .$$

Accumulators

More reliable sensitivity and specificity may be achieved by accumulating features of $\{TT_n\}$ or $\{TT_n - TT_{fi}\}$ than by mere comparison of a single value to a threshold. Various accumulators are used in *statistical process control* [6, chapter 5]. We re-label the sequences of interest to include their time stamp t: $\{TT_{n,t}\}$ and $\{TT_{n,t} - TT_{fi,t}\}$. Any of the accumulators below may yield satisfactory performance:

- *Runs:* Alarm if R successive values are over a threshold K.
- *Duration:* Alarm if all values within some time range are over a threshold K.
- *Cumulative sum (CUSUM):* Alarm if the accumulated sum of values over a threshold K exceeds another threshold M.
- *Exponentially weighted moving average (EWMA):* Alarm if a geometrically weighted sum of W_t exceeds a threshold K:

$$W_t = r\left(TT_{n,t} - TT_{fi,t}\right) + (1 - r) W_{t-1} \ ; \ W_0 = 0;$$

r between 0 and 1.

Just as when using threshold tests for significant trends, one should test any of these accumulation techniques on training data to choose parameter estimates yielding acceptable sensitivity and specificity.

TRIGGERING ALTERNATE MONITOR SETS

Monitor sets represent the set of competing trends in a diagnostic context. The significance of a fault trend may be a sign that the diagnostic context has changed. This in turn may warrant triggering of an alternate monitor set.

The monitor set representation may be supplemented to include rules of the following form:

If fault trend template TT_f is significant, then trigger monitor set M_j. The temporal distance between point P of TT_f and the anchor point Q on each trend template of M_j is the range $[t_b, t_e]$.

The significance test may be determined via any of the means in the previous section. TrenDx can apply these rules to monitor trends in a changing diagnostic context. Then each of the trend hypotheses in the new context are temporally anchored to process the new data within the appropriate intervals.

GENERATING ALARMS

Having information that a fault trend is significant at time T_0, an automated monitor can send an alarm to operators or clinicians caring for the process. The alarm can state the name of the trend template and the time T_0. Additional text may include descriptions of important value constraints or intervals. The text for these trend template components come fairly naturally from the knowledge representation and from the qualitative temporal interval relations. For example, the text for an alarm of a handbagging hemodynamic fault may read:

Handbagging hemodynamic fault was detected at 12:27:21 a.m. A phase of decreasing blood pressure proceeded a phase of increasing heart rate.

An alarm may also display some or all of the data in the fault intervals for the best matching fault hypothesis at time T_0.

An automated monitor may wish to alarm based on a boolean combination of significant trends. Forward chaining of a set of production rules may follow any of the significance methods of the previous section.

A more thorough method for deciding whether to send an alarm at time T_0 is evaluation of a *decision model* weighing the costs and benefits of sending the alarm versus not sending the alarm. The utility of alarming is based on:

- estimated probabilities of the normal and abnormal trends, and
- probabilities, costs, and benefits of each action a clinician may take upon seeing the alarm.

The utility of not alarming is based on the trend probabilities as well as

- probabilities, costs, and benefits of each action a clinician may take upon not seeing the alarm.

Considered as a single decision to made at time T_0, one can encode this decision model straightforwardly as a decision tree. This decision is more accurately made in the context of the time progression of the patient, partially as reflected by the best matching trend templates at each time. A critical review of *dynamic decision modeling* techniques is in [7].

DATA VISUALIZATION

Operators in monitoring environments with multiple channels of high frequency data may have extreme difficulty tracking all of this data for significant trends. An intelligent trend detector such as TrenDx may be used to intelligently filter this data to show the operator only that data corresponding to an important com-

ponent of a faulty process. This display can be driven by rules of the following form:

> If fault trend template TT$_f$ is significant, then display data of parameter P$_i$ during the time specification T$_i$ (for i from 1 to some integer k).

The significance test may be determined as previously discussed. Rules may be specific to the monitored process and the operator observing the data.

Each time specification T$_i$ is a Boolean combination of temporal intervals, and may be expressed in a temporal query language such as the time-line language of [8]. Intervals of trend templates can be the time intervals over which data should be displayed.
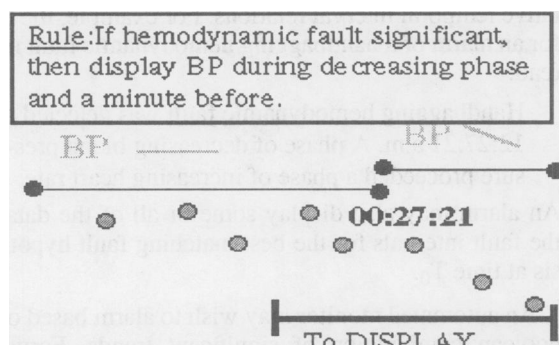

Figure 5: Filtering a data stream using TrenDx results.

An example intelligent display appears in Figure 5. The rule states that when the trend for a handbagging hemodynamic fault is significant, the display should show the blood pressures assigned to the decreasing phase of the trend template and a minute before. The optimal hypothesis for this fault trend is used for sending some data to the display. Thus this high frequency data stream has been filtered. This filtering technique is part of a broader data visualization effort [9].

RELATED WORK

Our work is compatible with time series analysis and signal processing [10]. When applied without context these techniques may produce high false positive rates and redundant alarms. These should lessen if such methods are used within trend template value constraints, which provide an appropriate temporal context.

Others have investigated monitoring using knowledge-based temporal patterns. Keravnou and Washbrook [11] have built a temporal model for diagnosing skeletal dysplasias. Their representation is more limited than trend templates in that time points cannot be fully flexible, and symptoms are limited to tokens for qualitative states. Chemical engineers [12, 13] have used qualitative descriptions of temporal trends in terms of first and second derivatives of single variables. Using rules with temporal patterns in antecedents, they have had promising results monitoring data streams similar to an

ICU. These representations are limited in only handling univariate trends over fixed time slices.

REFERENCES

[1] Haimowitz, I. J. and I. S. Kohane. "Automated Trend Detection with Multiple Temporal Hypotheses." IJCAI-93, Chambery, France, 146-151, 1993.

[2] Haimowitz, I. J. and I. S. Kohane. "An Epistemology for Clinically Significant Trends." AAAI-93, Washington, DC, 176-181, 1993.

[3] Nichols, D. G., J. J. McCloskey and M. C. Rogers. "Adult Respiratory Distress Syndrome." In *Textbook of Pediatric Intensive Care*. Baltimore, MD, Williams and Wilkins, 1993.

[4] Haimowitz, I. J. *Knowledge-Based Trend detection and Diagnosis*. Doctoral thesis, Massachusetts Institute of Technology, May 1994.

[5] Sachs, L. *Applied Statistics*. New York, Springer-Verlag, 1984.

[6] Ryan, T. P. *Statistical Methods for Quality Improvement*. Wiley Series in Probability and Mathematical Statistics. New York, Wiley, 1989.

[7] Leong, T.-Y. "Dynamic Decision Modeling in Medicine: A Critique of Existing Formalisms," SCAMC-93, Washington, D.C., 478-484, 1993.

[8] Cousins, S. B. and M. G. Kahn "The Visual Display of Temporal Information." *Artificial Intelligence in Medicine*, 3: 341-357, 1993.

[9] Fackler, J. and Kohane, I. "Hypothesis-Driven Data Visualization: SmartDisplay," SCAMC-94.

[10] Priestly, M. B. *Spectral Analysis and Time Series*. New York, Academic Press, 1981.

[11] Keravnou, E. T. and J. Washbrook. "A Temporal Reasoning Framework Used in the Diagnosis of Skeletal Dysplasias." *Artificial Intelligence in Medicine*, 2: 239-265, 1990.

[12] Cheung, J. T.-Y. and G. Stephanopolous. "Representation of Process Trends - Part II. The Problem of Scale and Qualitative Scaling." *Computers in Chemical Engineering*, 14(4): 511-539, 1990.

[13] Konstantinov, K. B. and T. Yoshida. "Real-Time Qualitative Analysis of the temporal Shapes of (Bio)process Variables." *AIChE Journal*, 38(11): 1703-1715, 1992.